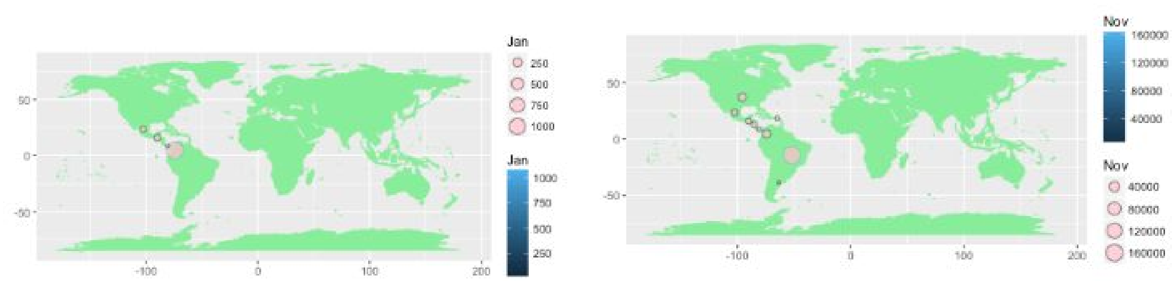
Mapping the Spread of Zika across the Americas: A Regional Approach
-Amna Aftab, Bennice Bao, Vera Zhang, Sukanya Sravasti, Swexya Panta

The Zika virus is a mosquito-borne virus. It is primarily spread through infected mosquitoes and can also be transmitted through sexual contact. The virus itself causes relatively mild symptoms that last for about a week. Most adults do not seek treatment and are not diagnosed. There is no known cure for the Zika virus. Historically, cases have been limited to Africa and Asia. The Zika virus has recently become the focus of an ongoing public health emergency as a result of a 2015 outbreak in Brazil which has led to widespread panic throughout the Americas. The rise in the spread of the Zika virus has been accompanied by an increase in cases of microcephaly. Microcephaly is a severe brain malformation where an infant's head is approximately two standard deviations below the norm. Children with microcephaly have reduced life expectancies and debilitating symptoms, including intellectual disability, speech delay and seizures. Research suggests that pregnant women infected with the virus can pass on the virus to their babies, which causes microcephaly and other brain defects.

We framed our approach around investigating the factors that affected the spread of the Zika virus in 2016. Specifically, we chose to look at demographic factors. We started off with some exploratory data analysis on the Zika outbreak on a global scale before focusing on the heart of the outbreak in Brazil. We then focused on USA and Panama. We paid particular attention to the effect that temperature changes over time and population density had over the spread of the cases. Our analysis suggests that population density, age demographics, travel and climate conditions are possible drivers for the spread of zika virus.

Mapping the worldwide outbreak: Climate changes



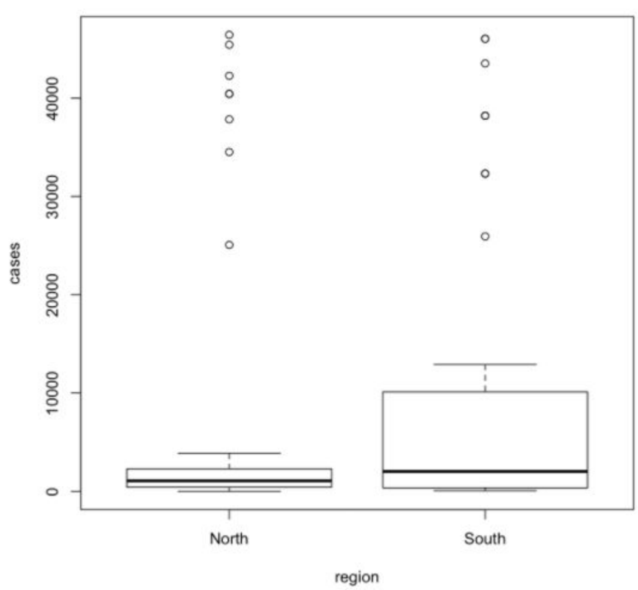
We examined the outbreak of zika cases on a global scale over the course of 2016 (specifically from January to November). From the graphs above, we found some interesting trends. Initially, we noted that zika was mostly reported in Central American countries. Over time, the number of cases reported from North and South America increase - particularly in the US and Brazil. There appears to be a greater outbreak of zika cases from April onwards. Countries near the equator experience a rainy season around this time and humidity can reach 70%. It is also well established that mosquitoes tend to thrive in humid conditions. We can then theorize that the increase in humidity might be correlated to the increase in zika cases. Complete list of graphs for particular months are given in the appendix.

Zika in Brazil: North versus South Population Density

As a result of the 2015 outbreak of the Zika virus in Brazil, there has been an ongoing public health emergency in Brazil. Since Brazil is one of the Latin American countries with a large number of reported Zika cases, we analyzed the possible causes behind the outbreak of Zika in Brazil.

Using data from the Centers For Disease Control and Prevention, we explore the trends in reported Zika cases across different Brazilian states. The data is collected for different months in 2016 at frequent intervals. Firstly, we categorize the Brazilian states into North and South based on their location in Brazil; states in northern Brazil are categorized as North and states in southern Brazil are categorized as South. States of northern Brazil include: Rondonia, Acre, Amazonas, Roraima, Para, Amapa, Tocantis, Maranhao, Piaui, Ceara, Rio Grande do Norte, Paraiba, Pernambuco, Alagoas, Sergipe, Bahia. States of southern Brazil include: Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina and Rio Grande do Sul.

The mean number of zika cases in states in the northern region is 3642 whereas the mean in southern region states is 8043. This leads us to carry out a hypothesis test where we test if region has an impact on number of zika cases. The boxplots below shows the distribution of zika cases in northern versus southern Brazil. As seen in the boxplots, the median and upper quartile for reported zika cases is greater in southern states compared to northern states.



We estimate the following model using OLS:

$$Zika\ Cases\ =\ \beta 1 * Region\ +\ \ e$$

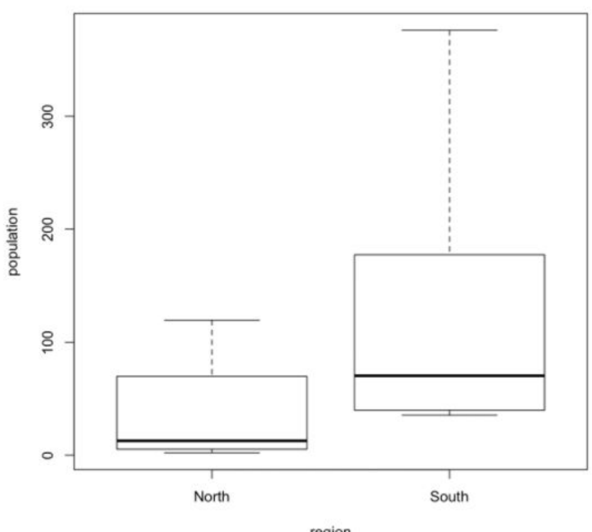
The results are summarized in the table below:

Variable	**Coefficient**	**Standard Error**
Region (South)	4401***	1698

***Indicates significance at the 5% level.

The dependent variable is the number of reported zika cases in 2016 and the independent variable is region. Region is a dummy variable indicating whether the state is in the northern (0) or southern region (1). The results show that the coefficient for the southern state is positive and significant. This means that if the state is located in the southern region, then the model estimates there are 4401 more reported cases of zika. However, the coefficient above could be biased due to other demographic factors of southern states that might lead to higher cases of zika. This might lead to omitted variable bias.

The results above led us to consider some other demographic factors that led to high number of zika cases in southern states compared to northern states. We obtained data on state-wise population density for the different states from the Brazilian Official Territorial Area website. The distribution of population density is shown in the boxplots below. As seen in the box plots, the population density for southern states is much higher compared to the northern states. As a result we add in population density to our previous equation.



We estimate the following model using OLS:

$$Zika\ Cases\ =\ \beta 1 * Region\ + \beta 2 * Population\ Density +\ \ e$$

The dependent variable is the number of reported zika cases and the independent variables are the Region and Population Density. Population density is the population density of the particular state in people per kilometers squared. The results of our regression are summarized in the table below:

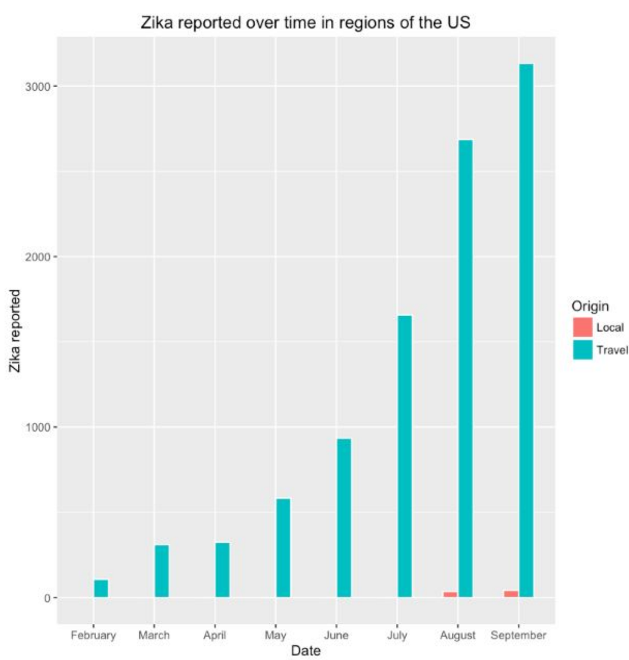
Variable	Coefficient	Standard Error
Region	-1951	1736
Population Density	74***	9

***indicates significance at the 5% level

The coefficient for population density is significant at the 5% level. This means that a 1 unit increase in population density leads to 74 more cases of zika reported. This makes sense since zika is transmitted through sexual contact and having a greater population density leads to higher chances of people interacting with each other, which in turn could increase the spread of zika. Including the population density variable removed the bias in the previous regression, so that the coefficient for region is no longer significant.

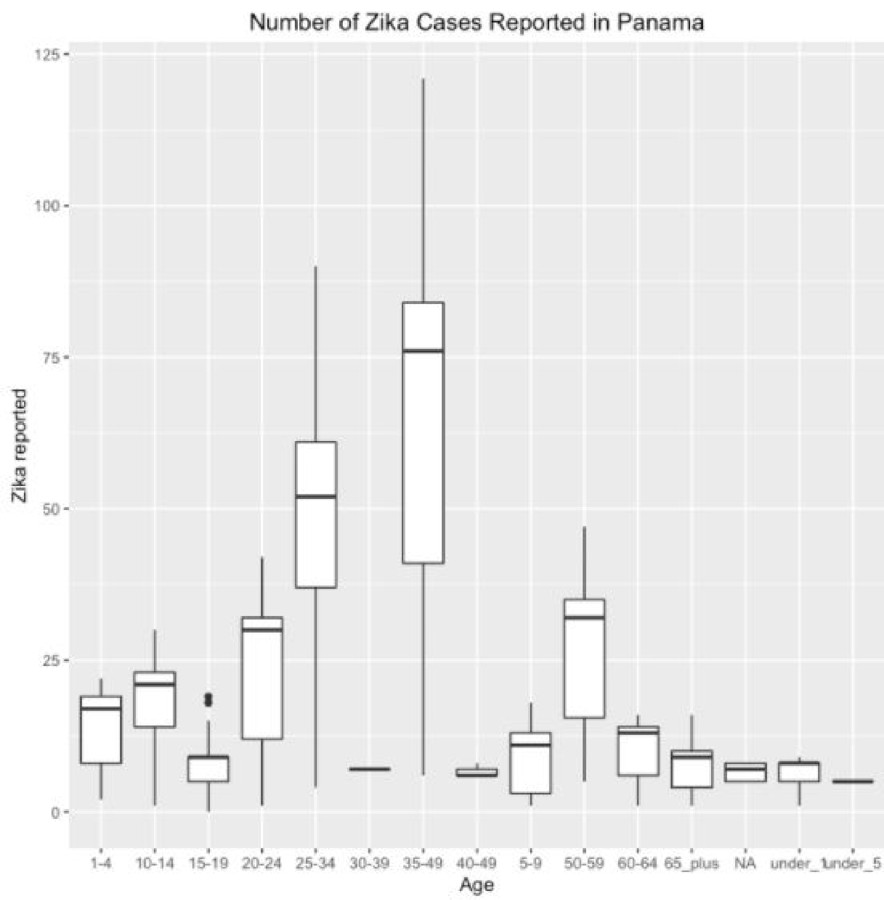
Assessing the Spread of Zika in the US: Travel and Climate

We also looked at cases of Zika in the United States. These cases were reported in areas where mosquitoes did not normally breed, which suggests that the virus was spread through travel. The figure below shows the spread of Zika over 2016. We can see that the number of travel cases rose over time, with the first set of local cases seen in August. This analysis in in line with our initial hypothesis that the disease is most likely to spread in warmer months.



Zika in Panama: Age Demographics

The graph below shows the distribution of reported zika cases in Panama. This suggests that that there are certain age groups that account for a greater percentage of reported zika cases than others. The age groups 25-34 and 35-49 have the highest number of reported zika cases. We theorize that this might be because these age groups are more socially active and travel more compared to other age groups. Hence, they are in contact with a greater number of people, which might increase their chances of being infected. In particular, these age groups could also be more sexually active, which might cause them to be infected with zika since the virus is spread through sexual contact.



Limitations and Conclusion:

One of the major problems that we faced during this study was the lack of available data. For example, we hypothesized that there might be a relation between the availability of healthcare services and the number of reported cases. Unfortunately the lack of statewide health care services data available for Brazil was a major barrier to us carrying out that study.

Another limitation to our results is that the data only accounts for the number of reported cases of the Zika virus. We suspect that certain parts of the world, particularly the poorer areas which might lack the resources to do accurate reporting, might not be reporting the true Zika cases. Also, the zika cases mentioned in the data are reported and are not confirmed, which causes a measurement error in our dependent variable. This is a limitation since this does not necessarily represent the true spread of the zika virus.

For future research, as more healthcare data becomes available for Brazil, we could analyze whether states with greater healthcare services are more likely to manage a zika epidemic with greater efficiency compared to states with poorer health care services.

Sources:

1. Center for Disease Control: https://www.cdc.gov/
2. Thu, Hlaing Myat, Khin Mar Aye, and Soe Thein. "The effect of temperature and humidity on dengue virus propagation in Aedes aegypti mosquitos." (1998).
3. Brazilian Official Territorial Area: http://www.ibge.gov.br/home/geociencias/areaterritorial/principal.shtm